GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2002, 23:33:11 ; Search time 1527.5 Seconds
                                          (without alignments)
                                          16168.980 Million cell updates/sec

Title: US-10-025-514-7
Perfect score: 1525
Sequence: 1 tctagaccatgtctggaaag.........ccaactcagaagtagtcgac 1525

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Listing first 45 summaries

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Database :

EST:*
    1: em_estba:*
    2: em_esthum:*
    3: em_estin:*
    4: em_estmu:*
    5: em_estov:*
    6: em_estpl:*
    7: em_estro:*
    8: em_htc:*
    9: em_est1:*
   10: gb_est2:*
   11: gb_htc:*
   12: gb_est3:*
   13: gb_est4:*
   14: gb_est5:*
   15: em_estfun:*
   16: em_estom:*
   17: gb_gss:*
   18: em_gss_hum:*
   19: em_gss_inv:*
   20: em_gss_pln:*
   21: em_gss_vrt:*
   22: em_gss_fun:*
   23: em_gss_mam:*
   24: em_gss_mus:*
   25: em_gss_other:*
   26: em_gss_pro:*
   27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	433.2	28.4	2478	11	AF130068	AF130068 Homo sapi
2	431.6	28.3	2571	11	AF113676	AF113676 Homo sapi
3	331.6	21.7	1392	11	AK002537	AK002537 Mus muscu
4	318.8	20.9	1296	11	AK004999	AK004999 Mus muscu
5	309.8	20.3	887	11	BQ643710	BQ643710 AGENCOURT
6	305.4	20.0	907	14	BQ648909	BQ648909 AGENCOURT
7	289.6	19.0	895	14	BQ653587	BQ653587 AGENCOURT
8	287	18.8	985	14	BM924019	BM924019 AGENCOURT
9	285.2	18.7	982	14	BQ646142	BQ646142 AGENCOURT
10	283.8	18.6	924	14	BQ958958	BQ958958 AGENCOURT
11	274.8	18.0	1194	14	BM924813	BM924813 AGENCOURT
12	273.4	17.9	907	14	BQ650189	BQ650189 AGENCOURT
13	271.4	17.8	891	14	BQ646948	BQ646948 AGENCOURT
14	270.4	17.7	959	14	BQ648524	BQ648524 AGENCOURT
15	268	17.6	796	12	BG567260	BG567260 60258974l
16 c	264.8	17.4	833	14	BQ007663	BQ007663 UI-H-EIO-
17	264.8	17.4	991	14	BQ064738	BQ064738 AGENCOURT
18	263.8	17.3	917	14	BQ650502	BQ650502 AGENCOURT
19	263.6	17.3	687	10	AV649258	AV649258 AV649258
20	263	17.2	916	14	BQ644130	BQ644130 AGENCOURT
21	262	17.2	686	12	BG569467	BG569467 60258892７
22 c	261	17.1	1012	14	BQ064473	BQ064473 AGENCOURT
23	260.2	17.1	977	14	BQ654015	BQ654015 AGENCOURT
24	259.6	17.0	740	10	AV653778	AV653778 AV653778
25	259.6	17.0	944	14	BQ645533	BQ645533 AGENCOURT
26	256	16.8	869	13	BI219058	BI219058 602938673
27	254	16.7	917	14	BG650698	BG650698 AGENCOURT
28	253	16.6	654	12	BG563537	BG563537 602581944
29	251.4	16.5	884	12	BG536548	BG536548 602564956
30	250.2	16.4	746	12	BG566873	BG566873 602589565
31	249.8	16.4	745	9	AI114643	AI114643 HA1237 Hu
32	249.6	16.4	1044	9	AI528087	AI528087 u193c09.y
33	249.6	16.4	813	13	BI759000	BI759000 603042479
34	249.4	16.4	764	12	BF384318	BF384318 602046666
35	249	16.3	731	10	AV649144	AV649144 AV649144
36	248.8	16.3	804	12	BG618951	BG618951 602646361
37 c	247.4	16.2	813	14	BQ182052	BQ182052 UI-H-EUO-
38 c	245.8	16.1	780	9	AI110706	AI110706 HA0145 Hu
39	244.2	16.0	992	13	BI246204	BI246204 602958716
40	243.8	16.0	752	12	BG618001	BG618001 602644877
41	243.2	15.9	674	14	BM724546	BM724546 UI-E-EO1-
42	243.2	15.9	759	12	BG533293	BG533293 601860666
43	243	15.9	749	9	AI526727	AI526727 u142b09.y
44	242.8	15.9	790	13	BI761275	BI761275 603044028
45	242.6	15.9	934	14	BQ646994	BQ646994 AGENCOURT

ALIGNMENTS

RESULT 1
AF130068
LOCUS   AF130068   2478 bp   mRNA   linear   HTC 08-MAY-2001
DEFINITION Homo sapiens clone FLB8226 PRO2209 mRNA, complete cds.
ACCESSION   AF130068
VERSION   AF130068.1 GI:11493442
KEYWORDS   HTC.
SOURCE   Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2478)
AUTHORS   Zhang,C., Yu,Y., Zhang,S., Wei,H., Bi,J., Zhou,G., Dong,C., Zal,Y.,
            Xu,W., Gao,F., Liu,M. and He,F.
TITLE   Functional prediction of the coding sequences of 75 new genes
            deduced by analysis of cDNA clones from human fetal liver
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2478)
AUTHORS   Zhang,C., Yu,Y., Zhang,S., Wei,H., Bi,J., Zhou,G., Dong,C., Zal,Y.,
            Xu,W., Gao,F., Liu,M. and He,F.
TITLE   Direct Submission
JOURNAL   Submitted (23-FEB-1999) Department of Experimental Hematology,
            Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing,
            Beijing 100850, P. R. China
FEATURES   Location/Qualifiers
     source   1..2478
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="FLB8226"

```
CDS             /tissue_type="liver"
                /dev_stage="fetus"
                1082..2338
                /note="predicted protein of HQ2209"
                /codon_start=1
                /evidence=not_experimental
                /product="PRO2209"
                /protein_id="AAG35496.1"
                /db_xref="GI:11493443"
                /translation="MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHHDQDH
                PFFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILE
                GLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKK
                LYHSEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKW
                ERPEVKDTEEEDFHVDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIF
                FLPDEGKLQHLENELTHDIITKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITIK
                VFSNGADLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAAGAMFLEAIPMSIPPEVKFNK
                PFVFLMIEQNTKSPLFMGKVVNPTQK"

BASE COUNT      650 a    627 c    597 g    604 t
ORIGIN

Query Match            28.4%;       Score 433.2;  DB 11;   Length 2478;
Best Local Similarity  59.7%;       Pred. No. 4.9e-101;
Matches  729;  Conservative    0;  Mismatches  493;  Indels    0;  Gaps     0;

Qy    298 TGTGTGTAAGTCCTGTGTTCCCAGTCAAGGCCATGAAGAGACCCTCAAGGCGACGCCG 357
              ||  | ||| ||  | |||| |||||||||||||||||||||| |||||||||||||||
Db   1116 TGGCAGGCCGTGTGCCCTGGTCCCTGTCTCCCTGGCTGAGGATCCCCAGGGAGATGCTG 1175

Qy    358 CTCAAAAAACGACACCAGTCATCACGACGACCAAGAGACCATCCGACTTTAATAAAATTACTC 417
              |||||   | ||| ||||||||| |||||||| ||| ||| |||||||||||||||||
Db   1176 CCCAGAAGACAGATACATCCCACCATGATCAGGATCAACCAAACCTTCAACAAGATCACCC 1235

Qy    418 CAAATTTAGCCGAATTGCTTTTCTTTGTATAGACAATTAGCTCATCAAAGTAATTCTA 477
              ||||||||||||||||||||| |||||||  |||||||| || |||||||||||||||
Db   1236 CCAACCTGGCTGAGTTCGCCTTCAGCCTAGCCGGTGGCACCAGTCCAACAGCA 1295

Qy    478 CTTAACATTTTTTTTAGTCCTGTTTCATTGCCACTGCTTTCGCCATGTAGTTTAGTA 537
              ||  | ||| ||||||||| |||||| | |||||| ||||| |||||||| | |||||
Db   1296 CCAATATCTTCTTCTCCCCAGTGAGCATCGGTCACAGCCTTTGCAATGCTCTCCCCTGGGGA 1355

Qy    538 CTAAAGCCGATACCCATGACGAGATTTTAGAGAGTTTAAACTTAATTGACCGAAATCC 597
              ||||||| | || |||| |||||||| |||||||  | ||| ||||| | |||||| |
Db   1356 CCAAGGCTGACACTCACGATGAAATCCTGGAGGGCCTGAATTCAACCTCACGGAGATTC 1415

Qy    598 CAGAGAGCCCAAATTCCACGAGGGTTTTCAAGAGTTGTTGAGAACTTTGAATCAACCTGAT 657
              |||||||||||||||| |||||||| ||||||| ||||| || | |||||||| || |||
Db   1416 CGGAGGCCTCAGATCCATGAAGGCTTCCAAGAACTCCTCCGTACCCTGAACCAGCCAGACA 1475

Qy    658 CTCAATTGCAATTAACTACTGGTAACGGTTTATTTTGTCTGAAGGTTAAAATTGGTTG 717
              |||||| || ||||| ||| || ||| | |||||  || ||||||||| ||||| |||
Db   1476 GCCAGCTCCAGTCGACCACGGCCAATGGCCTGTTCCTCAGCGAGGGCCTGAAGCTAGTGG 1535

Qy    718 ACAAATTCCTAGAGACGTCAAGAAACTATATCATAGTGAGGCTTTACGGTTAATTTTG 777
              ||||||||| |||| | ||||||  |||| | ||||| ||||| | ||| |||||| ||
Db   1536 ATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACTCAGAAGCCTTCACTGTCAACTTCG 1595

Qy    778 GTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTAATGTTGAGAAAGGCACCCAGGGTA 837
              ||||| |||||||| |||||||| ||||||||||||| ||||||||||||||||||| ||
Db   1596 GGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAAGGGTACTCCAGGCA 1655

Qy    838 AGATCGTTGACCTAGTTAAAGAATTAGATCGTGATACCGTCTTCGCACTAGTTAACTATA 897
              || | || |||| |||| ||| ||||||  | ||||  |||||| | ||||| ||||||
Db   1656 AAATTGTGGATTTGGTCAAGGAGCTTGACAGAGACACAGTTTTTGCTCTGGTGAATTACA 1715

Qy    898 TTTTTTTCAAGGTAAGTGGGAACGTCCTTTGCGAGGTTAAAGATACTGAAGAGGAAGATT 957
              | | || || || | | ||||| |  ||| |||||| | |||| || ||||||||||| |
Db   1716 TCTTCTTTAAGGCAATAGGCAATGCCACCGCTATTTTCTTCCTTCCAGACGAGGGAAAGCT 1775

Qy    958 TTCATGTTGATCAAGTTACTACTGTCAAGTTCCAATGATGAAGAAGACTGGGTATGTTCA 1017
              ||||||| || |||| ||| | ||||||||||| || |||| ||||| |||||| | |||
Db   1776 TCCAGTGGACAAGGTTCGACACCGTGAAGGTGCCTATGATGAAGCCTTTAGGCATGTTTA 1835

Qy   1018 ATATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTAATGAACTATTTAGGTAACG 1077
```

```
RESULT 2
AF113676
LOCUS       AF113676        2571 bp    mRNA    linear   HTC  08-MAY-2001
DEFINITION  Homo sapiens clone FLB2803 PRO0684 mRNA, complete cds.
ACCESSION   AF113676
VERSION     AF113676.1  GI:6855600
KEYWORDS    HTC.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2571)
  AUTHORS   Zhang,C., Yu,Y., Zhang,S., Ouyang,S., Luo,L., Wei,H., Zhou,G.,
            Zhou,W., Bi,J., Zhang,Y., Liu,M. and He,F.
  TITLE     Functional prediction of the coding sequences of 32 new genes
            deduced by analysis of cDNA clones from human fetal liver
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2571)
  AUTHORS   Zhang,C., Yu,Y., Zhang,S., Ouyang,S., Luo,L., Wei,H., Zhou,G.,
            Zhou,W., Bi,J., Zhang,Y., Liu,M. and He,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1998) Department of Experimental Hematology,
            Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing
            100850, P. R. China
FEATURES             Location/Qualifiers
     source          1..2571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="FLB2803"
                     /tissue_type="liver"
                     /dev_stage="fetus"
                     /note="HQ0684"
     repeat_region   515..819
                     /rpt_family="Alu"
     CDS             1109..2365
                     /note="predicted protein of HQ684"
                     /codon_start=1
```

```
/evidence=not_experimental
/product="PRO0684"
/protein_id="AAF29581.1"
/db_xref="GI:6855601"
/translation="MPSSVSWGILLLAGLCLVPVSLAEDPQGDAAQKTDTSHHDQDH
GLNFNLIEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKK
LYHSEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKW
ERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIF
FLPDEGKLQHLENELTHDIIITKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITK
VFSNGADLSGVTEEAPLKLSRAVHKAVLTIDEKGTEAAGAMFLEAIPMSIPPEVKFNK
PFVFLMEQNTKSPLFMGKVVNPTQK"
```

BASE COUNT   800 a   638 c   556 g   577 t
ORIGIN

Query Match   28.3%;   Score 431.6;   DB 11;   Length 2571;
Best Local Similarity   59.6%;   Pred. No. 1.3e-100;
Matches   728;   Conservative   0;   Mismatches   494;   Indels   0;   Gaps   0;

```
Qy    298  TGTGTGGTAAGTCCTGTGTTCTCCCAGTCAAGGCCATGGAAGACCCTCAAGGCGACGCCG  357
                || |||  ||| ||||| ||| |||||||||| |||||| || ||||
Db   1143  TGGCAGGCCTGTGTCGTCGGTCCTGTCCCCTGGCTGAGGATCCCCAGGGAGATGCTG    1202

Qy    358  CTCAAAAAACCGACACCAGTCATCACGACCAAGACCATCCGACTTTTAATAAAATTACTC  417
                || ||||| ||| ||| |||| || |||| |||| || |||| ||
Db   1203  CCCAGAAGACACAGATACATCCACCAATGATCAGGATCACCCAACCTTCAACAAGATCACCC  1262

Qy    418  CAAATTTAGCCGAATTGCTTTTCTTTGTATAGACAATTAGCTCATCAAAGTAATTCTA   477
                || ||| |||||| ||  ||||| ||||| ||| |||| ||||  ||||||
Db   1263  CCAACCTGGCTGAGTTCGCCTTCAGCCATCAGCCACCAGCTGCACACCAGTCCAACAGCA  1322

Qy    478  CTAACACATTTTTTTAGTTCCTGTTTCTATTGCCACTGCTTTCGCCAGTGTTGAGTTTAGGTA  537
                ||| ||||||     |||  ||| ||  ||| ||||| ||| ||  |||||
Db   1323  CCAATATCTTCTCTCCCCAGTGACGACGATTTGCAGGCATCGCTACACGAGCCTTTGCAG  1382

Qy    538  CTAAAGCCGATACCCATGACGAGATTTAGAAGGTTTAAACTTAAATTGACCGAAATCC    597
               | |||||| || ||| ||| ||  |||||||| ||  ||| |||||| |||
Db   1383  CCAAGGCTGACACTCACGATGAAATCCTGAGGGCCTGAATTCAACCTCACGGAGATTC  1442

Qy    598  CAGAAGCCCAAATTCACGAGGGTTTCAAGAGTGTTGTTGAGAACTTTGAATCAACCTGATT  657
                |||||| ||||| ||||   |||||  || | ||| ||||| ||||
Db   1443  CGGAGGCTCAGATCCATGAAGGCTTCCAGGAACTCCCGTACCCTCAAACCAGCCAGACA  1502

Qy    658  CTCAATTGCAATTAACTACTGGTAACGGTTTATTTTTGTCTTGAAGTTTAAAATTGGTTG  717
                ||| ||  ||| || ||| ||||| ||||||||||| ||||  ||| ||||
Db   1503  GCCAGCTCCAGCTGACCACCGGCAATGGCCTGTTCCTCAGCGAGGGCCTGAAGCTAGTGG  1562

Qy    718  ACAAATTCCTAGAAGCGTCAAGAAACTATATCATAGTGAGGCGTTTTACCGTTAATTTTG  777
                | ||| ||||| || ||| ||| || ||| ||||| ||| |||  ||| || ||
Db   1563  ATAAGTTTTGGAGGATGTTAAAAGTGTTACCACTCAGAAGCCTTCACTGTCAACTTCG  1622

Qy    778  GTGATACTGAGGAAGCCAAGAAGGCTAAAAAGCAAATTAATGATGATGTGTGAGAAAGGCACCCAGGGTA  837
               ||  ||  ||| || |||||| || || |||| ||| ||| || ||| ||
Db   1623  GGGACACCGAAGAGGCCAAGAAGCAGATCAACGATTACGTGGAGAAGGGTACTCAAGGGA  1682

Qy    838  AGATCGTTGACCTAGTTAAAGAATTAGAGATCGTGATACCGTCTTCCGACTAGTTAACTATA  897
              | |||||| || ||| |||  ||| || |||||| || |||| ||||  ||||  ||
Db   1683  AAATTGTGGATTTGGTCAAGGAGCTTGACAGAGACACCGTTTTTGCTCTGGTGAATTACA  1742

Qy    898  TTTTTTTCAAGGGTAAGTGGGAACGTCCTTTCGAGGTTAAAGATACTGAAGAGGAAGATT  957
                ||||||||||| ||||| |||||||| ||| |||| ||||| ||| |||||
Db   1743  TCTTCTTTAAGGGAAATGGGAAGTGGGAGAGACCCTTTGAAGTCAAGGACACCGAGGAAGAGGACT  1802

Qy    958  TTCATGTTGATCAAGTTACTACTGTGTCAAGTTCAATGATGAAGAAAGACTCGGGTATGTTCA  1017
                |||  || ||||| || ||| |||| ||  ||| |||| |||| ||||
Db   1803  TCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTA  1862

Qy   1018  ATATTCAACATTGCAAAAAAATTAAGTTCTTGGGTCTTATTAATGAACGTATTTAGTTAAGTATTTAGGTAACG  1077
               |  ||| |||| ||||| ||| |||| || || |||  |||| ||  || |||||| ||
Db   1863  ACATCCAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCTAATGAAATACCTGGGCAATG  1922

Qy   1078  CTACTGCTATTTTTTTTTTTACCAGCGAAGGTAAGCTTCAACATTTAGAGAAGTAGTTGA  1137
               | || || |||| |  ||  || | ||  || |||| ||| |  ||  |  ||  |||
```

RESULT 3
AK002537
LOCUS   AK002537   1392 bp   mRNA   linear   HTC 19-JAN-2002
DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:0610011G14:serine protease inhibitor 1-4, full
            insert sequence.
ACCESSION   AK002537
VERSION   AK002537.1   GI:12832592
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:0610011G14.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,S., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto.H., Sakaguchi.S., Ikegami,T., Kashiwagi.K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu.M., Inoue.Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861

```
Db   1923  CCACCGCCATCTTCTTCTGCCTGATGAGGGAAACTACAGCACCTGGAAAATGAACTCA  1982
                || ||| | ||||  || ||||  ||| |||| ||| ||||||  || |
Qy   1138  CTCATGACATATTACTAAATTTTAGAACGAGGATCGTCGTAGCGGCTTCTCTGCCACC  1197
                |||  ||| ||  |||| |||  ||| ||||||| |  ||  || |||
Db   1983  CCCACGATATCATCACCCAAGTTCCTGGAAATGAAGACAGAAGGTCTGCCAGCTTACATT  2042
                ||| | ||  ||  |||||  || |||| ||||||  || |||||| ||
Qy   1198  TGCCAAAGTTAAGTACCGGTACTTACGACTAAAATCTGTTTTAGGCCAGTTAGGTA  1257
                ||||| ||||||||| ||||| |||| ||||| |||||||| ||||||
Db   2043  TACCCAAACTGTCCATTACTGGAACCTATGATCTGAAGAGCGTCCTGGGTCAACTCGGCA  2102
                ||| ||||| |  ||||||  | ||| ||  ||||| || ||| ||||
Qy   1258  TTACCAAAGTTTTTCTTCAACGGTGCCGATTTGAGTGTGTTACTGAAGAAGTCCATTAA  1317
               || ||||||  | | ||| ||| ||| |||| || |||| ||| ||| ||||||
Db   2103  TCACTAAGGTCTTCAGCAATGGGGCTGACCTCTCCGGGTCACAGAGGAGGCACCCCTGA  2162
                || |||| || |||||  | ||||| || |||  |||| || || |||||
Qy   1318  AATTGAGTAAAGCTGTTCACAAAGCCGTCTTAACTATTGATGAAAATTAATA  1377
               || ||| || |||||| ||||||  || ||||||| || || |||||
Db   2163  AGCTCTCCAAGGCGTGCATAAGGCTGTGTCTGACCATCATGGACGAGAAAGGGACTGAAGCTG  2222
                || || || || ||  || |  ||| ||  ||| ||||||| ||| ||||
Qy   1378  CCGGGCCTATGTCCTGGAAGCTATTCCAATGAGCATTCCACCAGAAGTTAAATTTAATA  1437
               | |||  ||||||  |||| || ||||| || | ||||||||  |||||||
Db   2223  CTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCGAGGTCAAGTTCAACA  2282
                ||||| |||||| || ||| ||| ||||| ||| ||| |||||  ||||||
Qy   1438  AACCATTCGTTTTCTGATGATCGAGCAGAGAACACTAAAAGCCATTGTTTATGGGTAAGG  1497
               ||||  ||||| ||| || || |||  | |||| | |||||  ||||||||  |||
Db   2283  AACCCTTTGTCTTCTTAATGATTGAACAAAAATACCAAGTCTCCCCTCTTCATGGGAAAAG  2342
               |||   |  | ||  || ||||| || ||||| || |||||  |||  |||||||
Qy   1498  TTGTCAACCCAACTCAGAAGTA  1519
               |||  |||  |||| ||||||
Db   2343  TGGTGAATCCCCACCAAAAATA  2364
```

REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5 (bases 1 to 1392)
AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
             Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
             Carninci,P., Fukuda,S., Fuknishi,Y., Furuno,M., Hanagaki,T.,
             Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
             Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
             Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
             Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
             Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
             Saito,R., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
             Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
             Tanaka,T., Tejima,Y., Toya,T., Yamamura,I., Yamanaka,I.,
             Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
             Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
             further details.
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues. First strand cDNA was primed with a primer
             [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse transcriptase
             and subsequently enriched for full-length by cap-trapper. cDNA went
             through one round of normalization to Rot = 5.0. Second strand cDNA
             was prepared with the primer adapter of sequence[5'
             GAGAGAGAGAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCC 3']. cDNA was
             cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
             XhoI. Host: SOLR.
FEATURES   Location/Qualifiers
     source   1..1392
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:0610011G14"
                     /db_xref="MGD:MGI:1896776"
                     /db_xref="taxon:10090"
                     /clone="0610011G14"
                     /sex="male"
                     /tissue_type="kidney"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"

     gene   /dev_stage="adult"
                     1..1392
                     /gene="Spi1-4"
     CDS   68..1309
                     /gene="Spi1-4"
                     /note="data source:MGD, source key:MGI:891968,
                     evidence:ISS
                     putative
                     serine protease inhibitor 1-4"
                     /codon_start=1
                     /protein_id="BAB22173.1"
                     /db_xref="GI:12832593"
                     /db_xref="MGD:MGI:891968"
                     /translation="MTPSISWSLLLAGLCLVPSFLAEDVQETDTSQKDQSPASHEI
                     ATNLGDFALRLYRELVHQSNTSNIFFSPVSIATAFAMLSLGSKGDTHTQILEGLQFNL
                     TQTSEADIHKSFQHLLQTLNRPDSELQLSTGNGLFVNNDLKLVEKFLEEAKNHYQAEV
                     FSVNFAESEAKKVINDFVEKGTQGKIVEAVKKLDQDTVFALANYILFKGWKKQPFDP
                     ENEEAAEFHVDESTTVKVPMMLSGMLDVHHCSMLSSWVLLMDYAGNTTAVFLLPDDG
                     KMQHLEQTLNKELISQFLLNRRSDAQIHIPRLSISGNYNKTLMSPLGITRIFNNGA
                     DLSGITEENAPLKLSKAVHKAVLTIDETGTEAAAATVLQVATYSMPPIVRFDHPFLFI
                     IFEEHTQSPIFVGKVVDFTHK"

BASE COUNT   364 a   393 c   324 g   311 t
ORIGIN
Query Match   21.7%;   Score 331.6; DB 11; Length 1392;
Best Local Similarity   56.3%;   Pred. No. 8.2e-75;
Matches 663; Conservative   0; Mismatches   509; Indels   6; Gaps   2;

Qy   344 TCAAGGCGACGCCGCTCAAAAAACCGACACCAGTCATCACGACCAAGACCATCCGACTTT 403
              || | ||| | |||||| || || |||| || ||| ||| | ||| ||||| |||
Db   133 TCTCGGCTGAGGATGTTCAGGAGACAGACACCTCCAGAAGGATCAGTCC--CCAGCCTC 189

Qy   404 TAATAAAATTACTCCAAATTAGCCGAATTTGCTTTTCTTGTATAGACAATTAGCTCA 463
              || |||| ||| || || ||| || || ||| | || || || |||| || ||
Db   190 CCATGAGATCGCTACAAACCTGGAGACTTTGCCCTCAGACTATACAGGGAGCTGGTCCA 249

Qy   464 TCAAAGTAATTCTACTAACATTTTTTTTAGTCCTGTGTTTCTATTGCCACTGCTTTCGCCAT 523
              ||| ||| ||| ||| | |||| || |||| | |||||| |||| || ||
Db   250 TCAGTCCAACACTTCAACATCTTCTTCTCCCCAGTGAGCATTGCCACAGCCCTTTGCTAT 309

Qy   524 GTTGAGTTTAGGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTAAACTTTAA 583
              | ||| |||| || |||| | | ||| || || |||| |||| | | ||
Db   310 GCTCTCCCTAGGGAGCAAGGGTGACCACTCACACCGAGATCCTGAGGGGCCTGCAGTTCAA 369

Qy   584 TTTGACCGAAATCCCAGAAGCCCAAATTCACGAGGGTTTCAAGAGTTGTTGAGAACTTT 643
              ||| | | || ||||||| || |||| || |||| ||||| | |||||||| || ||
Db   370 CCTCACACAAACATCGGAGGCTGACATCCACAAGTCCTTCCAACACCTCCTCCAAACCCT 429

Qy   644 GAATCAACCTGATTCTCAATTGCATTAACTACTGGTAACGGTTTATTTTGTCTGAAGG 703
              || | ||| | ||| || | ||||| || || | ||||| | ||||| |||| |
Db   430 CAACAGACCAGACAGTGAGCTGCAGCTGACGCACAGGCAATGGCCTCTTTGTCAACAATGA 489

Qy   704 TTTAAAATTGGTTGACAAATTCCTAGAAGACGTCAAGAAACTATATCAGTGAGGCTTT 763
              ||| | |||| |||| | ||||| || || |||| | || ||| || ||| | ||
Db   490 CCTGAAGCTGGTGGAGAAGTTTCTGGAGGAGGCCAAGAACCATTATCAGGCAGAAGTCTT 549

Qy   764 TACCGTTAATTTGGTGATACTGAGGAAGCTAAAAAGCAAATAATGATTAATGTTGAGAA 823
              | | ||| || || | || ||||| |||| | |||| || | | ||||||||| |
Db   550 CTCTGTCAACTTTGCAGAGTCAGAGGCAGGCCAAGAAAAGTGATTAATGATTTTGTGGAGAA 609

Qy   824 AGGCACCCAGGGTAAGATCGGTAACGTCCAGTTAAGAATTAGATCGTGATACCGTCTTCGC 883
              | | | ||| | || | | |||| | | || ||||| | ||| || || | ||| ||
Db   610 GGGAACCCAGGAAGAAAGATAGTTGAGGCGGTGAAAAAACTGGACCAAGACACAGTTTCGC 669

Qy   884 ACTAGTTAACTATATTTTTTTCAAGGTAAGTGGGAACGTCCTTTCGAGGTTAAAGATAC 943
              |||| ||||| | |||||| ||||| | || || | || | || || | |||
Db   670 CCTGGCGAATTACATTCTCTTTAAAGGCAAATGGAAAACAGCCATTCGATCCTGAGAACAC 729

Qy   944 TGAAGAGGAAGATTTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATGAAAAG 1003
              | ||| || |||| ||| ||| || || | || | || | |||| | |
Db   730 TGAGGAGCTGAGTTCCATGTGACGAGTCGCCACCAGTGAAGGTGCCCATGATGATGACCCT 789

Qy   1004 ACTGGGTATGTCAATATTCAATATTCAAAAAATTAAGTTCTTCGGTGTTATTAATGAA 1063

```
Db    790 CTCGGGCATGCTTGATGTGCATCATTGCAGCATGCTGTCCAGCTGGGTGCTCGCTGATGGA 849
              |||  |||  ||  |||  ||  |||||||     |  |||    |  |  ||  |  |  ||||
Qy   1064 GTATTTAGGTAACGCTACTGCTATTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTT 1123
              ||   || |||||||  ||   |||||||| ||||  |||||||||||||||||  |||
Db    850 TTACGCAGGCAACACCACTGCTGTCTTCCTCCCTGCCCGATGATGGGAAGATGCAGCATCT 909
Qy   1124 AGAGAATGAGTTGACTCATGACATTATTACTAAATTTTTAGAGACGAGGATCGTCGTAG 1183
              |||  |     |  ||  |  |    |  |||  |  |||||||||  | |  ||||
Db    910 GGAGCAAACTCTCAACAAGGAGCTCATCTCCTCAGTTCCTGCTAAACAGGCCGCAGAAGCGA 969
Qy   1184 CGCTTCTCTGCACCTGCCAAAGTTAAGTATCACCGGTACTTACGACTTAAAATCTGTTTT 1243
              ||| | |  || |||||  |||  || ||  |  |||| ||| || |||||| || ||||
Db    970 TGCTCAGATCCATATCCCCAGACTGTCCATCTCTGGAAACTATAACTTGAAGACACTCAT 1029
Qy   1244 AGGCCAGTTAGGTATTACCAAAGTTTTTTCTAACGGTGCCGATTGAGTGGTGTTACTGA 1303
              |||  ||| |  || ||  |||||||||   | |||||| |||  ||| || ||| |||
Db   1030 GAGTCCACTGGGCATCACCCGGATCTTCAACAATGGGGCTGACCTCTCCGGAATCACAGA 1089
Qy   1304 AG---AAGCTCCATTAAAATTGAGTAAAGCTGTTCACAAAGCCGTCTTAACTATTGATGA 1360
              ||    |||  || ||  ||| ||  |  ||| || | ||  || ||  | ||||||||
Db   1090 GGAGAATGCTCCCCTGAAGCTCACAAGGCTGTGCATAAGGCTGTGCGACCATCGATGA 1149
Qy   1361 AAAGGGTACCGAGGCGCCGGCGCCCTATGTTCCTGGAAGCTAATTCCAATGAGCATTCCACC 1420
              ||  |  |||  || ||  || || |  | |  || ||  |||| |||  |  | |||||||
Db   1150 GCACGGAACAGAGAAGCTGCAGCAGCTACAGTCTTTACAAGTCGCTACTTATTCTATGCCCCC 1209
Qy   1421 AGAAGTTAAAATTTAATAAACCATTCGTTTTCTGATGATGCGAGCAGAACACTAAAAGCCC 1480
              ||  |||  || || |  |  |  ||   |||  |||||||  || |||||| ||||| |||
Db   1210 TATCGTCCGTTCGACCACCGTTTCCTTTCCATTCATATATTGAAGAACACACTCAGAGCCC 1269
Qy   1481 ATTGTTTATGGGTAAGGTTGTCAACCCAACTCAGAAGT 1518
              |   ||  |  || ||  ||  |   |||  ||||| |||
Db   1270 CATCTTTCTGGGAAAAGTGGTAGAATCCCACCACACATAAAT 1307
```

RESULT 4
AK004999

LOCUS   AK004999   1296 bp   mRNA   linear   HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300014A17:serine protease inhibitor 1-4, full
            insert sequence.
ACCESSION   AK004999
VERSION   AK004999.1 GI:12835611
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1300014A17.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Toyoda,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,K., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
            and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5 (bases 1 to 1296)
AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Kasukawa,T., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,A.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAAGGATCCAAGCTCAATTAATTAATTAAACCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES   Location/Qualifiers
     source   1..1296
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1300014A17"
                     /db_xref="MGD:MGI:1889235"
                     /db_xref="taxon:10090"
```

```
              /clone="1300014A17"
              /sex="male"
              /tissue_type="liver"
              /clone_lib="RIKEN full-length enriched mouse cDNA library"
              /dev_stage="adult"
     gene     1..1296
              /gene="Spi1-4"
     CDS      48..1289
              /gene="Spi1-4"
              /note="data source:MGD, source key:MGI:891968,
              evidence:ISS
              putative
              serine protease inhibitor 1-4"
              /codon_start=1
              /protein_id="BAB23733.1"
              /db_xref="GI:12836612"
              /db_xref="MGD:MGI:891968"
              /translation="MTPSISWCLLLLAGLCCLVPSFLAEDVQEfDTSQKDQSPASHEI
              ATNLGDFAISLYRELVHQSNTSNIFFSPVSIATAFAMLSLGSKGDTHTQLEGLQFNL
              TQTSRADIHNSFQHLLQTLNRPDSELQLSTGNGLFVNNDLKLVEKFLEERAKNHYQAEV
              FSVNFAESEEAKKVINDFVEKGTQGKIVEAVKKLEQDTVFVLANYILFKGKWKKPFDP
              ENTKQAEFHVDSETTVKVPMMTLSGMLDVHHCSTLSSWVLMLDYAGNATAVFLLPDDG
              KMQHLEQTLNKELISKFLLNRRRLAQIHIPRLSISGNYNLETLMSPLGITRIFNSGA
              DLSGITEENAPLKLSQAVHKAVLTIDETGTEAAAATVLQGGFLSMPPILHFNRPLFI
              IFEEHSQSPLFVGKVVDPTHK"
BASE COUNT    339 a    370 c    300 g    287 t
ORIGIN

Query Match          20.9%;
Best Local Similarity 55.6%;
Matches   655;    Conservative   0;    Mismatches   517;    Indels   6;    Gaps   2;

Qy   344 TCAAGGCGACGCCGCCTCAAAAAACCGACACCAGTCATCACGACCAAGACCATCCGACTTT 403
         |||| || ||| ||  ||| | |||| || ||| ||  |||  ||  ||||| ||||| |
Db   113 TCTGGCTGGAGGATGTTCAGGAGACAGACACCTCACCACACTCCCAGAAGGATCAGTCC---CCAGCCTC 169

Qy   404 TAATAAAATTACTCCAAATTAGCCGAATTGCTTTCTTTGTATAGACAATTAGCTCA 463
         | ||| ||| | ||| | |||| ||  ||| ||  |||| ||
Db   170 CCATGAGATTGCTACAAACCTGGGAGACTTTGCCATCAGTCATACCGGGAGCTGGTCCA 229

Qy   464 TCAAAGTAATTCTACTAACATTTTTTTTAGTCCTGTTTCTATTGCCACTGCTTTCGCCAT 523
         | | ||| || |||| || |||| | || || ||| | || ||| ||||||
Db   230 TCAGTCCAACACTTCCAACATCTTCTCTCCCCAGTGAGCATTGCCACAGCCTTTGCTAT 289

Qy   524 GTTGAGTTAGGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTAAACTTTAA 583
         || ||| | ||| ||| || |||||| || ||  ||| | |||||| |||| |||||
Db   290 GCTCTCCCTAGGGAGCAAGGGTGACACTTCACACGACAGATCCTAGAGGGCCTGCAGTTCAA 349

Qy   584 TTTGACCGAAATCCCAGAAGCCCAAATTCACGAGGGTTTTCAAGAGTTGTTGAGAACTTT 643
         | ||| ||| ||| |||| || | |||| |||| || |||| ||| || ||
Db   350 CCTCACAAACATCGGAGGCTGACATCCACAATTCCTTCCAACACCTCCTCCAAACCCT 409

Qy   644 GAATCAACCTGATTCTCAATTGCAATTAACTACTGGTAACGGTTTATTTTGTCTGAAGG 703
         || || || ||||| || |||| |||| |||| ||| || ||| | ||| ||||
Db   410 CAACAGACCAGCCAGTGAGCTGCAGCGTGCAGCCAATGGCCGTCTTGTCAACAATGA 469

Qy   704 TTTAAAATTGGTTGCACAATTCCTAGAAGACGTCAAGAAACTATATCAGTGAGGCTTT 763
         || |||| | ||| |||| || ||| ||| ||| || |||| | |||| |||||
Db   470 CCTGAAGCTGGTGGAGAAGTTTCTGGAGGAGCGAGCTAAAAAGCCAATTATCAGGCAGAAGTCTT 529

Qy   764 TACCGTTAATTTGGTGATACTGAGGAAGTGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAA 823
         | ||| || ||| ||| | |||| |||| || || |||| ||| | |||| |||
Db   530 CTCTGTCACTTTCGCAGAGTCAGAGGAGGCCAAGAAAGTGATTAATGATTTTGTGGAGAA 589

Qy   824 AGGCACCCAGGTAAGATCGTTGACCTAGTTAAAGAATTAGATCGTGATACCGTCTTCGC 883
         || || ||| |||| || |||| | ||| |||| ||| || | ||| ||||| |
Db   590 GGGAACCCAAGGAAAGATAGTTGAGGCTGTGAAAAAACTAGAACAAGACACAGTTTTCGT 649

Qy   884 ACTAGTTAACTATATTTTTTCAGGTAAGGTAAGTGGGAACCTCCTTTCAGAGGTTAAAGATAC 943
         |||| || ||| | |||| | |||| |||| || ||| ||  | ||| ||||| | ||
Db   650 CCTGGCAAATTACATTCTCTTTAAAGGCAAATGGAAGAGAGCCATTGACCTGAGAACAC 709

Qy   944 TGAAGAGGAAGATTTTCATGTTGATCAAGTTACTACTGTTCAAGTTCCAATGATGAAAAG 1003
```

```
Db   710 TAAGCAAGCTGAGTTCCACGTGGACGAGTCCACCACGGTGAAGGTGCCCAGATGATGACCCT 769
         || || || |||| |||| | || ||| | || ||| || || |||| ||| ||||
Qy  1004 ACTGGGTATGTTCAATATTCAAGATTGCAAAAAATTAAGTTCTTGGTTCTTATTAATGAA 1063
         ||| | | || | ||||  |||| | |||||
Db   770 CTCGGGCATGCTGACGTGCCACCATTGCAGCACGCTCCCAGCTGGTGCTGCTGATGGA 829

Qy  1064 GTATTTAGGTAACGCTACTGCTATTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTT 1123
         || ||| || || |||| || || |||||| | | || || ||| ||| || |||||
Db   830 TTACGCGGGCAACGCACTGCTGTCTTCCTGCCCGATGAGGGAAGATGGCAGCATCT 889

Qy  1124 AGAGAATGAGTGACTCATGACATTATTACTAAATTTTAGAGAACGAGGATCGTCGTAG 1183
         |||| | || || ||| |||| | |||| |||| || |||| | || |||| || ||
Db   890 GGAGCAAACTCAACAAGGAGTCCATCTTAAGTTCCTGCTAAACAGGCCAGAAGGTT 949

Qy  1184 CGCTTCTCTGCCACCTGCCAAAGTTAAGTATCACCGTACTTACGACTTAAAATCTGTTTT 1243
         || || | | ||| |||| || |||| ||  |||| | ||||| || |||| || |
Db   950 AGCCCAGATCCATATCCCCAGCGTCCATCTCTGAAACATAACTTGGAGACACTCAT 1009

Qy  1244 AGGCCAGTTAGGTATTACCAAAGTTTTTCAACGGTGCCGATTTGAGTGGTGTTACTGA 1303
         |||| || ||| | ||||  ||| |||| |||| || || | |||| || || |||
Db  1010 GAGTCCACTGGGCATCACCCGGATCTTCAACAGTGGGGCTGACCTCTCCGGAATCACAGA 1069

Qy  1304 AG---AAGCTCCATTAAAATTGAGTAAAGCTGTTCACAAAGCCGTCTTAACTATTGATGA 1360
         ||   |||| | ||| ||  | ||| || || |||| || |||| | || |||| |||
Db  1070 GGAGAATGCTCCCCTGAAGCTGAGCCAGGCTGTGCATAAGGCCGTGCTGACCATCGATGA 1129

Qy  1361 AAAGGGTACCGAGGCCGCGGCCCATGTTCCTGGAAGCTATTCCAATGAGCATTCCACC 1420
         |||| |  ||| | |||| ||  |||| || || |||| |||| || | ||| |||||
Db  1130 GACAGGACAGGAAGCTGCAGCAGCTCCAGTGCTTACAGGGCGTTTTTGTCTATGCCCCC 1189

Qy  1421 AGAAGTTAAATTTAATAAACCATTCGTTTTCTGATGATCGAGCAACACTAAAGCCC 1480
         || | || | |  | |||| | || |||  | || |  | ||| | |||| ||||
Db  1190 TATCTTGCACTTCAACCGCCCTTCCTTTTCCTTTTCATAATATTTGAAGAACACTCTCAGAGCCCC 1249

Qy  1481 ATTGTTTATGGGTAAGGTTGTCAACCCAACTCAGAAGT 1518
         | ||| || || ||| || ||  |||| ||||| |
Db  1250 CCTCTTTTGTGGGAAAAGTGGTAGATCCCACAACATAAAT 1287


RESULT 5
BQ643710
LOCUS       BQ643710                 887 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8342217 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268225
            5', mRNA sequence.
ACCESSION   BQ643710
VERSION     BQ643710.1  GI:21767882
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 887)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
COMMENT     Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2439 row: n  column: 02
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1..887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6268225"
                     /clone_lib="NIH_MGC_100"
```

```
Qy  1359 GAAAAGGGTACCGAGGCGCCGCGGCGCTATGTCCTGGAAGCTATTCC 1405
         ||| || || |||||||||| ||| ||  ||||  |  |||||
Db   840 GAGAAAGGGACTGAAGCTGCTGGGCGCCCATGTTTTTAGAGGCCATACC 886


RESULT 6
BQ648909
LOCUS       BQ648909            907 bp    mRNA     linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8351269 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286839
            5', mRNA sequence.
ACCESSION   BQ648909
VERSION     BQ648909.1  GI:21773081
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 907)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2488 row: e  column: 16
            High quality sequence stop: 721.
FEATURES             Location/Qualifiers
     source          1..907
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6286839"
                     /clone_lib="NIH_MGC_100"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT    246 a    228 c    239 g    192 t    2 others
ORIGIN

Query Match             20.0%;    Score 305.4;  DB 14;  Length 907;
Best Local Similarity   59.8%;    Pred. No. 4.4e-68;
Matches   530;  Conservative   0;  Mismatches  356;  Indels   1;  Gaps   1;

Qy   541 AAGCCGATACCCATGACGAGATTTTAGAACTTTAATTTGACCGAAATCCCAG 600
         | ||| || || ||| || |  |||| ||| |||| ||||||| |
Db    15 AGGCTGACACTCACGATGAAATCCTGGAGGCCTGAATTTCAACCTCACGGAGATTCCGG 74

Qy   601 AAGCCCAAATTCACGAGGGTTTTCAAGAGTTGTTGAGAACTTTGAATCAACCTGATTCTC 660
         ||     |  ||  |||  |||||| |  |||| ||  ||  ||  |||||
Db    75 AGGCTCAGATCCATGAAGGCTTCCAGGAACTCCGTACCCAGCCAGACAGCC 134

Qy   661 AATTGCAATTAACTACTGGTAACGGTTTATTTTGTCTGAAGGTTAAAATTGGTTGACA 720
         | |||||| |||| ||  |||||||| ||  ||| |||||||| ||  ||||
Db   135 AGCTCCAGCTGACCACCGGCCAATGGCCTGTTCCTCAGCGAGGGCCTGAAGCTAGTGGATA 194

Qy   721 AATTCCTAGAGACGTCAAGAAACTATATCATAGTGAGGCTTTACCGTTAATTTTGGTG 780
         | || || ||  ||||| | || | |||||  |||| ||| ||  |||||||| |
Db   195 AGTTTTTGGAGGATGTTAAAAAGTTGTACCACTCAGAAGCCTTCACTGTCAACTTCGGGG 254
```

```
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT    242 a    224 c    232 g    187 t    2 others
ORIGIN

Query Match             20.3%;    Score 309.8;  DB 14;  Length 887;
Best Local Similarity   60.0%;    Pred. No. 3.1e-69;
Matches   532;  Conservative   0;  Mismatches  354;  Indels   1;  Gaps   1;

Qy   519 GCCATGTTGAGTTTAGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTTAAAC 578
         ||  ||  ||  |||| |  |||||| ||  ||  ||  |||  |||
Db     1 GCAATGCTCTCCCCGTGGGACCCAAGGCTGCACACTCACGATGAAATCCTGGAGGCCTGAAT 60

Qy   579 TTTAATTTGACCGAAATCCCAGAAGCCCAAATTCACGAGGGTTTTCAAGAGTTGTTGAGA 638
         |||||  | || |||||| ||  ||| |  |||  |||  |||||| |
Db    61 TTCAACCTCACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCCAGGAACTCCGTACCCGT 120

Qy   639 ACTTTGAATCAACCTGATTCTCCAATTGCAATTAACTACTGGTAACGGTTTATTTTTGTCT 698
         ||  |  | |||| |   |||| || ||| || ||||||  ||  ||||
Db   121 ACCCTCACCAGCCAGACCAGCTCCAGCTGACCACCGGCCAATGGCCTGTTCCTCCAGC 180

Qy   699 GAAGGTTAAAATTGGTTGACAAATTCCTAGAGACGTCAAGAAACTATATCATAGTGAG 758
         || || ||  ||| ||  |||| | | ||  | || ||  ||  |||||  ||||
Db   181 GAGGGCCTGAAGCTAGTGGATGGATGTTAAAAAGTTGTACCACTCAGAA 240

Qy   759 GCTTTTACCGTTAATTTTGGTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTT 818
         || || || || ||  || | | ||| ||  |  || |||| || ||  |||| ||  |
Db   241 GCCTTCACTGTCAACTTCGGGGACACCAGCGAGAGGCCCAAGAAACAGATCAACGATTACGTG 300

Qy   819 GAGAAAGGCACCCAGGGTAAGATCGTTGACCTAGTTAAAGAATTAGATGCTGTGATACCGTC 878
         ||| ||||| |  |||| |  ||  | ||   || ||  | |||| |  | ||  | |||||
Db   301 GAGAAGGGTACTCAAGGGAAAATTGTGGATTTGGTCAAGGAGCTTGACAGAGACACAGTT 360

Qy   879 TTCGCACTAGTTAACTATATTTTTTTCAAGGGTGGGAACGTCCTTTCGAGGTTAAA 938
         |||||   ||||||| || |  || |||||  || || |||||||  | ||||
Db   361 TTTGCTCTGGTGAATTACATCTTCTTTAAAGGCAAATGGGAGCCCTTTGAAGTCAAG 420

Qy   939 GATACTGAAGGAGAAGATTTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATG 998
         |||| | ||||| ||||  | | | || || ||||  ||||||||  |||| ||| ||
Db   421 GACACCCGAGGAAGAGGACTTCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATG 480

Qy   999 AAAAGACTGGGTATGTTCAAATATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTA 1058
         || || || || |||| ||  |  || |||     || |   |  ||||||  |||
Db   481 AAGCGGTTTAGGCATGTTAACATCACCGATGTTAACATCCAGCACCACTGTAAGAAGCTGTGCCAGCTGCTGCTG 540

Qy  1059 ATGAAGTATTTAGGTAACGCTACTGCTATTTTTTTTTTTTACCAGACGCAAGGTAAGCTTCAA 1118
         |||||  | || || ||||| || ||| ||  ||||  |  |||| |  ||| ||||| ||||
Db   541 ATGAAATACCTGGGCAATGCCACCGCCATCTTCTTCTCCGCCTGCGATGAGGGGAAACTACAG 600

Qy  1119 CATTTAGAGAATGAGTGACTGATGCATGACATTATCATTAAATTTTAGAGAACGAGGATCGT 1178
         || || || ||| ||  |  |  ||   | ||  |||  |  ||||| |||||  ||| | |
Db   601 CACCTGGAAAATGAACTCACCCACGATATCACCAACAGTTCCTGGAAAATGATGAAGACAGA 660

Qy  1179 CGTAGCGGCTTCTCTGCACCTGCCAAAGTTAAGTATCACCGGTACTTACGACTTAAAATCT 1238
         ||   | |||| || |  ||  || | || |||||    ||  ||  ||  || || |||
Db   661 AGGTCTGCCAGCTTACCATTACCCAAACTGTCCATTACTGGAAACTGTCCATTACTCGGAAGAGC 720

Qy  1239 GTTTTAGGCCAGTTAGGTATTACCAAAGTTTTTCTAACGGTGCCGATTTGAGTGGTGTT 1298
         ||   |  | ||| || ||  ||  ||||  ||   ||| || |  |  |||||| ||
Db   721 GTCCTGNGTCAACTGGGCATCACTAAGGTCTTCAGCAATGGGCTGACCTCTCCGGNGTC 780

Qy  1299 ACTGAAGAAGCTCCATTAAAATTGAGTAAAGCTGTTCACAAAGCCGTCTTAACTATTGAT 1358
         ||  || || || ||  | |  || | |||||  ||||  ||  ||||| | |||| |||
Db   781 ACAGAGGAGGCACCCCTGAGGCTCTCCAAGGC-GTGCATAAGGCTGTGCTGACCATCGAC 839
```

```
Qy   781 ATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGCACCCAGGTAAGA   840
         ||||| || || || || |   || ||||   |  | |  || ||||  | |||
Db   255 ACACCGAAGAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAGAGGGTACTCAAGGAAAA  314

Qy   841 TCGTTGACCCAGTTAAAGATTAGATCGTGATACCGTCTTCGCCACTAGTTAACTATATTT   900
         | || |  ||||| |||||| || ||  || ||  ||| ||||| ||  |   | | |
Db   315 TTGTGGATTTGGTCAAGGAGCTTGACAGACACAGTTTTTGCTCTGTGAATTACATCT     374

Qy   901 TTTTCAAGGTAAGTGGAACGTCCTTTCGAGGTTAAAGATACTGAAGAGGAAGATTTTC    960
         | || ||||  | ||  | ||| |||| |||| ||||   | |||||  ||||
Db   375 TCTTTAAAGGCAAATGGCAGAGACCCTTTGAAGTCAAGGAAGAAGAGGACTTCC       434

Qy   961 ATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATGAAAAGACTGGGTATGTTCAATA  1020
         | ||| ||| | ||  |||  | |||||| ||||||| |  ||  | |||  ||||  | |
Db   435 ACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATGAGCGTTTAGGCATGTTTAACA    494

Qy  1021 TTCAACATTGCAAAAAATTAAGTTCTTGGGTTCTTATTAATGAAGTATTTAGGTAACGCTA 1080
         | ||||| ||  | ||  | |||  ||||| | |  ||||||| |   |||| ||| |  |
Db   495 TCCAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCCTGTGATGAAATACCTGGCAATGCCA  554

Qy  1081 CTGCTATTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTTAGAGAATGAGTTGACTC  1140
         | |||| ||||| ||  |   || ||||  |  |||||| |||||||||||||  |  |||
Db   555 CCGCCATCTTCTTCTCCTGCCTGATGAGGGAAACTACAGCACCTGAAAATGAACTCACCC    614

Qy  1141 ATGACATTATTACTAAATTTTAGAGAACGAGGATCGTCGTAGCCGTTCTGCCACCTGC    1200
         | ||| || || |   ||  |  ||||||| | | | | ||||| |||||||| ||
Db   615 ACGATATCATCACCAGCCAGTTCCTGGAAAATGAAGACAGAGAAGGTCTGCCAGTTTTAC   674

Qy  1201 CAAAGTTAAGTATCACCGGTACTTACGACTTAAAATCTGTTTAGGCCAGTTAGGTATTA   1260
         | |||| |||| ||||| |  ||| ||||| ||| | |  | ||| |  | ||| || |
Db   675 CCAAACTGTCCATTACTGGAACCTATGATCTGAAGAGCGTCTGGGTCAACTGGGCATCA    734

Qy  1261 CCAAAGTTTTTTCTAACGGTGCCGATTTGAGTGGTGCTACTGAAGAAGCTCCATTAAAAT  1320
         | || || || |   | |||||| ||| || |  ||  |||||||||| |  |  ||| |
Db   735 CTAAGGTCTTCAGCAATGGGCTGACCTCTCCGGGTCACAGAGGAGGCACCCGTGAAGC    794

Qy  1321 TGAGTAAAGCTGTTCACAAAGCCGTCTTCTAACTATTGATGAAAAGGGTACCGAGGCCGCCG 1380
         |  ||||||| |  |  |||||  | ||| | ||| |||||| || |||| ||||||||||
Db   795 TCTCCAGGGC-GTGCATAAGGCGTGTGCTGCCAGTGTGCTGACGACAGAAGGGACTGAAGCTGCTG 853

Qy  1381 GCGCTATGTTCCTGGAAGCTATTCCAATGAGCATTCCACCAGAAGTT             1427
         | ||| | || |||||| ||||  ||||| |||| || |||| ||
Db   854 GGGCCATTGTTTTAGAAGGCATACCCATGTCTAATCCCCAGAGGT              900
```

RESULT 7
BQ653587

```
LOCUS       BQ653587                 895 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8349591 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284550
            5', mRNA sequence.
ACCESSION   BQ653587
VERSION     BQ653587.1  GI:21777759
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 895)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2482 row: f  column: 07
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1..895
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6284550"
                     /clone_lib="NIH_MGC_100"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      227 a    248 c    230 g    190 t
ORIGIN

Query Match             19.0%;  Score 289.6;  DB 14;  Length 895;
Best Local Similarity   59.3%;  Pred. No. 5.3e-64;
Matches  493;  Conservative    0;  Mismatches 339;  Indels    0;  Gaps    0;

Qy   298 TGTGTGGTAAGTCCTGTGTTCCCCAGTCAAGGCCATGGAAGACCCTCAAGGCGACGCCG   357
         || || ||  | | ||| ||||| ||||| ||||  | |||||| ||||| |||| |
Db    57 TGGCAGGCCTGTGCCTGCTGCCCTGTCTCCCTGAGGATCCCCAGGGAGATGCTG      116

Qy   358 CTCAAAAAACCGACACCAGTCATCACGACCAAGACCATCCGACTTTTAATAAAATTACTC  417
         || ||||| || ||  |  || |||| ||  || |||| |||| |  ||| |  |||  |
Db   117 CCCAGAAGACAGATACATCCCACCATGATCAGGATCACCAACAAGATCACCC         176

Qy   418 CAAATTTAGCCGAATTGCTTTCTTCTTTGTTATAGACAATTAGCTCATCAAAGTAATTCTA 477
         |||||| ||| ||  |||  |||  |  ||| | ||||  |||| |||| | | |||  | |
Db   177 CCAAGAACAGATACATCCCCACCATGATCAGGATCACCGCCACTATACCGCCAGTGCCACAGCA 236

Qy   478 CTAACATTTTTTAGTCCTGTTCTTATTGAGCAATGAGTTAGGTGAGTTTAGGTA       537
         || || |    ||  || ||  | | ||| |  |  |  | |  ||| |||  |
Db   237 CCAATATCTTCTTCTCCCCAGTGAGCACATCGGTGAATTTAAACTTTAATTTGACCGAAATCC 296

Qy   538 CTAAAGCCGATACCCATGACGAGATTTAGAGATTTCAAACTTGGGGCCTGAAATCC       597
         || ||||  ||||| ||  | | || | |  | |  ||||  || | | |||||||
Db   297 CCAAGGCTGACACTCACGATGAAATCCTGGAGGGCCGAATTTCAACCTCACGACGGAGATTC  356

Qy   598 CAGAAGCCCAAATTCACGAGGGTTTCAAGATGTTGTTGAGAACTTTGTGAATCAACCTGATT 657
         |||| ||  |||||| |  || |  ||| ||  | |  |||| |  | |||| | || |||
Db   357 CGGAGGCTCAGATGCTCCAGAGGCCTTCCAGGACTCCTCCAGGAACTCCGGTACCCAGCACA   416

Qy   658 CTCAATTGCAATTAACTACTGGTAACGCGTTATTTGTCTGAAGGTTAAAGTTAAAATTGGTTG 717
         ||| || | || |  |  |  || | |  ||  | | |  |||| | || |  |  |||||
Db   417 GCCAGCTCCAGCTGACCACCGGCAATGGCCGTCGTTCCTCAGCGAGGGCCTGAGCCGTAGTGG   476

Qy   718 ACAAATCCTAGAAGACGTCAAGAAACTATATCATAGTGAGCGTTAACTGTAACTCAGTCG   777
         | ||  ||| ||| | | |||  ||| || ||  || |    | | |  |||  |||| |
Db   477 ATAAGTTTTTGGAGGATGTTAAAAAGTTAACCACTGACAGTTGTACCACTCTGGTGAATTACA  536

Qy   778 GTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGCACCCAGGGTA   837
         | ||| |||||| || ||||||| || |||  | || | || |  | || ||||| ||
Db   537 GGGACACCGAAGAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAGAGGGTACTCAAGGGA  596

Qy   838 AGATCGTTGACCTAGTTAAAGATTAGATCGTGATACCGTCTTCGCCACTAGTTAACTATATA 897
         | | | || ||| ||||||| || || ||  | | | || | |  | ||||| | |||
Db   597 AAATTGTGGATTTGGTCAAGGAGCTTGACAGACACAGTTTTTGCTCTGTGAATTACA      656

Qy   898 TTTTTTCAAGGTAAGTGGAACGTCCTTTCGAGGTTAAAGATACTGAAGAGGAAGGAAGATT 957
         | || ||||||  | || ||  |||| | |||| ||||   | |||||||||||  |||
Db   657 TCTTTAAAGGCAAATGGCAGAGACCCTTTGAAGTCAAGGAAGAAGAGGACGACGAGGAGGACT 716

Qy   958 TTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATGAGCGTATGTGTTCA       1017
         || | ||| ||| | || || |  | ||| | |||||| | |  ||  |||   |
Db   717 TCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATGAGCGTTTAGGCATGTTTA    776

Qy  1018 ATATTCAACATTGCAAAAAATTAAGTTCTTGGGTTCTTATTAATGAAGTATTTAGGTAACG 1077
         | || || |  | ||| || || ||  ||||| | | ||||||||||  ||||
```

```
Db   777  ACATCCAGCACTGTAAGAGAGCTGTCCAGCTGGGTGCTGCTGATGAAATACCTGGGGCATG  836
Qy  1078  CTACTGCTATTTTTTTTTTTTACCAGACGGAGGTAAGCTTCAACATTTAGAGAA  1129
Db   837  CCACCGGCCATCTTCTTTCTTTCTGCCTGGATGAGGGAAACTACAGCACCTGGGAAA  888
```

RESULT 8
BM924019
LOCUS   BM924019   985 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION AGENCOURT_6709923 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760183
            5', mRNA sequence.
ACCESSION   BM924019
VERSION   BM924019.1 GI:19374398
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 985)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12806 row: m column: 16
            High quality sequence stop: 707.
FEATURES   Location/Qualifiers
     source   1..985
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5760183"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female. Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon cloning). Average insert size 1.4 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber (Invitrogen). Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT   258 a   270 c   242 g   213 t   2 others
ORIGIN

Query Match   18.8%; Score 287; DB 14; Length 985;
Best Local Similarity 59.5%;
Matches 519; Conservative   0; Mismatches 351; Indels   2; Gaps   2;
                                  Pred. No. 2.6e-63;

```
Qy   298  TGTGTGGTAAGTCCTGTGTTCCCCAGTCAAGGCCATGGAAGACCCTCAAGGCGACGCCG   357
Db    54  TGGCAGGGCCTGTGCCTGGGTCCCTGTCCCTGGCTGAGGATCCCCAGGGAGATGCTG    113
Qy   358  CTCAAAAAACCGACACCAGTCATCACGACACCATCCGACTTTTAATAAAATTACTC     417
Db   114  CCCAGAAGACAGATACATCCCACCATGATCAGGATCACCCAACCTTCAACAAGATCACCC  173
Qy   418  CAAATTTAGCCGAATTGCTTTTCTTTTGTATAGACAATTAGCTCATCAAAGTAATTCTA   477
Db   174  CCAACCTGGCTGAGTTCGCCTTCGGCCTTATACCGCGCTGGCACGCTGCCAACAGCA     233
Qy   478  CTAACATTTTTTAGTCCTGCTTGTTCTATTGCCACTGCTTTCGCCACTGTTTGAGTTTAGTA  537
```

```
Db   234  CCAATATCTTCTTCTCCCCAGTGAGCATCGCTACAGCCTTTGCAATGCTCTCCCTGGGGA   293
Qy   538  CTAAAGCCGATACCCATGACGAGATTTAGAAGTTTAAACTTTAATTGACCGAAATCC     597
Db   294  CCAAGGCTGCACACTCACGATGAAATCCTGAGGGCCTGAATTCAACCTCACGGAGATTC    353
Qy   598  CAGAAGCCCAAATTCACGAGGGTTTCAAGACTTGTTGAGAACTTTGATCAACCTGATT     657
Db   354  CGGAGGCTCAGATCCATGAAGGCTTCCGAGAACTCCTCCGTACCCTCAACCAGCCAGACA   413
Qy   658  CTCAATTGCAATTAACTACTGTAACGGTTTATTTTTGTCTGAAGGTTTAAAATTGGTTG    717
Db   414  GCCAGCTCCAGCTGACCACCGCAATGGCCTGTTCCTCAGCGAGGGCCTGAAGCTAGTGG    473
Qy   718  ACAAATTCCTAGAAGACGTCAAGAAACTATATCATAGTGAGGCCTTTACCGTTAATTTTG   777
Db   474  ATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACCTCAGAGAGCCTTCACTGTCAACTTCG 533
Qy   778  GTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGCACCCAGGGTA   837
Db   534  GGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAAGGGTACTCAAGGGA   593
Qy   838  AGATCGTTGACCTAGTTAAAGAATTAGATCGTGATACCGTCTTCCCACTAGTTAACTATA   897
Db   594  AAATTGTGGATTTGGTTCAAGGAGCTTGACAGAGACACAGTTTTTGCTCTGGTGAATTACA  653
Qy   898  TTTTTTTCAAGGGTAAGTGGAACGTCCTTTCGAGGTTAAAGATACTGAAGAGGAAGATT   957
Db   654  TCTTCTTTAAAGGCAAATGGGAGAGACCCTTTGAGTCAAGGACACCGAGGAAGAGGACT   713
Qy   958  TTCATGTTGATCAAGTTACTACTGTCAAGTTCCAATGATGAAAAGACTGGGTATGTTCA  1017
Db   714  TCCACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTA   773
Qy  1018  ATATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTTAATGAAGTATTAGGTAACG  1077
Db   774  ACATCCAGCACTGTAAGAGCTGTTCCAGCTGGGTGCTGCTGATGAAATACCTGGGCCAATG   833
Qy  1078  CTACTGCTAT-TTTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTTAGAGAATGAGTTG 1136
Db   834  CCACCGGCCATCTTCTTCTTCTCTTCNTGGCCTGATGAGGGGAAACTACAGGCACTGGAAAATGAACTC 893
Qy  1137  ACTCATGACATTATTACTAAATTTTTAGAGAA  1168
Db   894  AC-CACGATATCATCACAACAAGGTTCCTGGAAAA  924
```

RESULT 9
BQ646142
LOCUS   BQ646142   982 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION AGENCOURT_8492569 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6296341
            5', mRNA sequence.
ACCESSION   BQ646142
VERSION   BQ646142.1 GI:21770314
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2504 row: a   column: 14

```
Qy  1346 CTTAACTATTGATGAAAAGGTACCGAGCGCCGGCGCTATGTTCCTGGAAGCTATTCC 1405
         | ||  || || ||| | ||  ||| | ||||| | ||  ||||| ||
Db   721 GCTGACCATCGACGAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTACGAAGGCATACC 780

Qy  1406 AATGAGCATTCCACCAGAGTTAAATTAATAAACCATTCGTTTTCTGA 1455
         |||  || | ||  | | ||  | ||||| || ||  |||| ||
Db   781 CATGTCTATCCCCCCGAGGGCAGTTCAACCAAACCCTTTGTCTTCTTA 830
```

RESULT 10
LOCUS   BQ958958   924 bp   mRNA   linear   EST 21-AUG-2002
DEFINITION AGENCOURT_10034701 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483305
            5', mRNA sequence.
ACCESSION   BQ958958
VERSION   BQ958958.1 GI:22374436
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 924)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2667 row: k column: 18
            High quality sequence stop: 586.
FEATURES   Location/Qualifiers
     source   1..924
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6483305"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   251 a   253 c   226 g   193 t   1 others
ORIGIN

Query Match   18.6%;   Score 283.8;   DB 14;   Length 924;
Best Local Similarity   58.4%;   Pred. No. 1.7e-62;
Matches 513; Conservative   0; Mismatches 363; Indels   2; Gaps   1;

```
Qy   408 AAAATTACTCCAAATTTAGCCGAATTTGCTTTTCTTTGTTATAGACAATTAGCTCATCAA 467
         || ||  ||| || |  ||  ||  || | || | ||  ||| | || || ||
Db     2 AAGATCACCCCAACCTGCCTGAGTTCGCCTTCCGCCCTATACCCAGCTGGCACACCAG 61

Qy   468 AGTAATTCTACTAACATTTTTTAGTCTCGTTCTTAGTCCGTTCTTCATTGCCACTGCGTTG 527
         | || |  | || | | ||  || | ||| | | | || |||| | | || ||| ||
Db    62 TCCAACAGCACCAATATCTTCTTCTCCCCAGTGACGACGATTTTAGAAGTTTAAACTTTTG 121

Qy   528 AGTTTAGGTACTAAAGCCGATACCATGACAGCAGATGTTTCAACTTTCAACCTC 587
         | |||  || | || | | ||||  || ||| |  ||  || || ||
Db   122 TCCCTGGGACCAAGGCTGACACTCACCATGATAAATCCTGGAAATCCTGAATTTCAACCTC 181

Qy   588 ACCGAAAATCCCAGAGCCCAAATTCACCAGGGTTTTCAAGAGTTGTTGAGAACTTTGAAT 647
         | ||  | | | || |  || || || | || || || |
Db   588 ACCGAAAATCCCAGAGCCCAAATTCACCAGGGTTTTCAAGAGTTGTTGAGAACTTTGAAT 647
```

FEATURES   High quality sequence stop: 647.
     source   Location/Qualifiers
                     1..982
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6296341"
                     /clone_lib="NIH_MGC_100"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT   267 a   259 c   238 g   217 t   1 others
ORIGIN

Query Match   18.7%;   Score 285.2;   DB 14;   Length 982;
Best Local Similarity   59.8%;   Pred. No. 7.6e-63;
Matches 496; Conservative   0; Mismatches 333; Indels   1; Gaps   1;

```
Qy   627 GAGTGTTGAGAACTTGAATCAACTGATTCTCAATTGCAATTAACTACTGGTAACGGT 686
         || ||  | || |  ||  | |  || || ||| || || | ||| || || ||
Db     1 GAACTCCTCCGTACCCCCAACCAGACAGACGCTCCCAGTGACCACCGGCAATGGC 60

Qy   687 TTATTTTTGTCTGAAGGTTAAAATTGGTTGACAAATTCCTAGAAGACGTCAAGAAACTA 746
         | | ||  ||  ||||| || || | |  | |||  || |  |||| | | |
Db    61 CTGTTCCTCAGCGAGCGGCCTGAAGCTAGTGGATGTAGTAAGTTTTTGGAGGATGTTAAAAAGTTG 120

Qy   747 TATCATAGTGAGGCTTTTACCGTTAATTTGGTGATACTGAGGAAGCTAAAAAGCAAATT 806
         || | |  ||| | || ||| ||  | |  || | || || ||| || || ||| ||
Db   121 TACCACTCAGAAGCCTTCACTGTCCACTTCGGGGACCACCGAAGAGAGGCCAAGAAACAGATC 180

Qy   807 AATGATTATGTTGAGAAAGGCACCCCAGGTAAGATCGTTGACCTAGTTAAAGAATTAGAT 866
         || || || ||| || || ||  | || || || || ||  | || | |||| || ||
Db   181 AACGATTACGTGGAGAAGGGTACTCAAGGGAAAATTGTGGATTTGCTCAAGGAGCTTGAC 240

Qy   867 CGTGATACCGTCTTCGCCACTAGTTAACTATATATTTTTTCAAGGTAGTGGGAACGTCCT 926
         | |||||| ||  ||||| ||| || |  | |||  ||| ||  ||  || || || ||
Db   241 AGAGACACAGTTTTGCTCTGGTGAATTACATCTTCTTTAAAGGCAAATGGGAGAGACCC 300

Qy   927 TTCGAGGTTAAAGATACTGAAGAGGAAGATTTCATGTTGATCAAGTTACTACTGTCAAA 986
         || || || || ||  || || ||||  || ||  | |  |||  || || ||| |||
Db   301 TTTGAAGTCAAGGACACCAGGAGGAGGACGTTCCACGTGACCAGGTGACCACCGTGAAG 360

Qy   987 GTTCCAATGATGAAAAGACTGGGTATGTTCAATATTCAACATTGCAAAAAATTAAGTTCT 1046
         || || ||||||| | | ||| ||||| |||  ||| | | |||  || ||| || ||
Db   361 GTGCCTATGATGAAGCGTTTAGGCATGTTTAACATCCAAGCACTGTAAGAAGCTGTCCAGC 420

Qy  1047 TGGGTCTTATTAATGAAGTATTAGGTAACGCTACTGCTATTTTTTTTTTCCTGCGTGATGAG 1106
         |||||  | | ||||| |  || ||||| ||   ||  || || || | ||  || ||||
Db   421 TGGGTGCTTGCTGATGAAATACCTGGGCAATGCAATGACATTATTACTAAATTTTAGAG 480

Qy  1107 GGTAAGCTTCAACATTAGAGAATGAGTTGACTCATGACAATCATTTTAGAG 1166
         ||  || | | || | ||  || ||  ||| ||  || || ||  || ||
Db   481 GGGAAACTACAGCACCTGGAAAATGAACTCACCACCAGATATCATCACCAAGTTCCTGAA 540

Qy  1167 AACGAGGATCGTCGTAGCCGTTCTCTGCACCTGCCAAAGTTAAGTATCACCGGTACTTAC 1226
         || || |  |  |||  || | ||| ||||   ||  ||||  || |  | |  || ||
Db   541 AATGAAGACAGAAGGTCTGCCAGCTTACATTTACCCAAACTGTCCATTACTGGAACCTAT 600

Qy  1227 GACTTAAAATCTGTTTAGGCCAGTTAGTATTACCAAAGT-TTTTCTTAACGGTGCCGA 1285
         ||  | ||| | ||  |||  || | || | ||||| || |||| | |||| || ||
Db   601 GATCTGAAGAGCGTCGGTCACTGGCATCCATAAGGTCTTTCAGCAAGTGGGGGCTGA 660

Qy  1286 TTTGAGTGGTGTTACTGAAGAAGCTCCATTAAATTGAGTAAGGCTGTTCACAAAGCCGT 1345
         ||| ||| || || |  || || | | || || || ||| | |  || || || |||
Db   661 CCTCTCCGGGGTCACGAGGAAGCCGTGATAAGGCCGTGCTAAGGCTGT 720
```

```
FEATURES        High quality sequence stop: 662.
source          Location/Qualifiers
                1..1194
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5761267"
                /clone_lib="NIH_MGC_116"
                /lab_host="DH10B"
                /note="Organ: pooled colon, kidney, stomach; Vector:
                pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                source anonymous pool of 3 colons, age 26 yo male, 49 yo
                female, 71 yo male colon; 46 yo male kidney, and pool of 2
                stomachs, 62 yo male and 70 yo female. Library is
                oligo-dT primed and directionally cloned (EcoRV site is
                destroyed upon cloning). Average insert size 1.4 kb,
                insert size range 1-3 kb. Library is normalized and
                enriched for full-length clones and was constructed by C.
                Gruber (Invitrogen). Research Genetics tracking code
                023. Note: this is a NIH_MGC Library."

BASE COUNT      321 a      303 c      328 g      242 t
ORIGIN

Query Match         18.0%;      Score 274.8; DB 14; Length 1194;
Best Local Similarity   61.2%;   Pred. No. 4e-60;
Matches    444; Conservative    0; Mismatches   282; Indels    0; Gaps    0;

Qy    772 ATTTTGGTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGCACCC 831
          ||  || | |  |||| ||||| | |  ||| || || |  |||| || |||||||
Db      7 AATTTCCGGGATCCGAAGAGCCAAGAAACAGATCAACGATTACGTGGAGAAGGGTACTC 66

Qy    832 AGGGTAAGATCGTTGACCTAGTTAAAGATTAGATCGTGATACCGTCTTCGCCACTAGTTA 891
          ||| |||||| |||  ||||| || | || |||| || | ||| || ||| |||||  |
Db     67 AAGGGAAAATTGTGGATTTGGTCAAGGAGTCAAGGACACACAGTTTTTGCTCTGGTGA 126

Qy    892 ACTATATTTTTTCAAGGTAAGTGGGAACGTCCTTTCGAGGTTAAAGATACTGAAGAGG 951
          ||| | |  ||| |||  ||| || | |||  |||| ||||| | | || ||| ||
Db    127 ATTACATCTTCTTTAAAGGCAAATGGAGGAGACCCCTTTGAAGTCAAGGACACCGAGGAAG 186

Qy    952 AAGATTTCATGTTGATCAAGTTACTACTGTCAAAAATTAAGTTCTTGGGTCTTATTAATGAAGTAATTTAG 1071
          ||  | | | || |||||| | | | |||  ||| || ||  ||| ||||||  || ||  | ||||| |
Db    187 AGGACTTCCACGTGACCAGGTGACCACCGTGAAGGTGCCTATGATGAGGCTTTAGGCA 246

Qy   1012 TGTTCAATATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTAATGAAGTAATTTAG 1071
          ||||  | | || || |  | |  || ||||  | |    | | || |  || |||| |
Db    247 TGTTTAACATCCACGTGACCACTGTAAGAAGCTGTCCAGCTGGCTGGTGCTGATGAAATACCTGG 306

Qy   1072 GTAACGCTACTGCTGTTATTTTTTTTTTTACCAGACGAAGTAAGCTTCAACATTTAGAGAAATG 1131
          |   | | | | |  |||  ||||||||  |  | || |||| | ||| ||||||||
Db    307 GCAATGCCACCGCCATCTTCTTCCTGCCTGATGAGGGAAACTACAGCACCTGGAAAATG 366

Qy   1132 AGTTGACTCATGACATTATTACCAAATTTTTAGAGACAAGACTTCCTGGAAAATGAAGAGCAGCGAGGATCGTCGTAGCGCTTCTC 1191
          |   | | | || | || ||  |||||| |||| |||||| || | | |||| | | ||||  |||
Db    367 AACTCACCCACGATACATCACCAAGTTCCTGGAAAATGAAGACAGAAATGAAGAGCAGAAGGTCTGCCAGCT 426

Qy   1192 TGCACCTGCCAAAGTAAGTAAGTATCACCGGTACTTACGACTTAAAATCGTCGAGAGCGTCCTGGGTCAAC 1251
          || | | | || || ||  || ||  | |  |||  || |||  || | |||||
Db    427 TACATTTACCCAAACTGTCCATTACGTGATCTGAAGGCTGTGCTGACCATCGACGACGAAAAGGGACTG 486

Qy   1252 TAGGTATTACCAAAGTTTTTCTAACGGTGCCGATTTGAGTGGTGTTACTGAAGAAGCTC 1311
          || | |   | |||  ||||| |  | |   |||||| |   |||   || |||  |||
Db    487 TGGGCATCACTAAGGTCTTCAGCAATGGGGCTGACCTCTCCCGGGTCACAGAGGAGCAC 546

Qy   1312 CATTAAAATTGAGTAAGCTGTTCACAAAGCCATCTTCAACATTGATGAATGAAAAGGTACCG 1371
          |  | ||| | || ||| | | || |||  | | || | | ||||| |||| | |||
Db    547 CCCTGAAGCTCTCCAAGCCGTGCATAAGGCTGTGTGACCATCGACGAGAAAAGGGACTG 606

Qy   1372 AGGCCGCCCGGGCCTATGTTCCTGGAAGCTATTCCAACAGAAGTTAAAT 1431
          || |||  | || |  | || || | |  | | | ||  | |  || ||
Db    607 AAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCGAGTCAAGT 666

Qy   1432 TTAATAAACCATTCGTTTTCTGATGATCGAGCAGAACACTAAAAGCCCATTGTTTATGG 1491
          || | |  | |  ||   | ||  || || |||  |||| | |  ||  | | ||| ||
Db    667 TCAACAAACCCTTTGTCTTCTTAATGATTGAACAAAATACCAAGTCCCCTCCCTTCATGG 726
```

```
Db    182 ACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCCAGGAACTCCTCCGTACCCTCAAC 241
Qy    648 CAACCTGATTCTCAATTGCAATTACTACTGGTAACGGTTTATTTTGTCTGAAGGTTTA 707
          ||| | |   ||    || ||  |  | | ||||  |  ||| ||  || |||| | |
Db    242 CAGCCAGACACAGCCAGCTCCAGCTGACCACCGGCCAATGCCGTGTTCCTCAGCGAGGGGCCTG 301
Qy    708 AAATTGGTTGACAAATTCCTAGAAGACGTCAAGAAACTATATCATAGTGAGGCTTTTACC 767
          ||| ||   ||  || ||  |  | |  |   |||  | || || || |  ||  ||
Db    302 AAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACTCAGAAGCCTTCACT 361
Qy    768 GTTAATTTGTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGC 827
          || |  |  ||  | | || ||| | |  ||| || |  || || | ||||| ||
Db    362 GTCAACTTCGGGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAAGGGT 421
Qy    828 ACCCAGGGTAAGATCGTTGACCTAGTTAAAGATTAGATCGTGATACCGTCTTCGCCACTA 887
          ||| || |  |||||| ||||| || | || |||  | || || |||  | ||| | |
Db    422 ACTCAAGGGAAAATTGTGGATTTGGTCAAGGAGTCAAGGACACACAGTTTTTGCTCTG 481
Qy    888 GTTAACTATATTTTTTTCAAGGTAAGTGGGAACGTCCTTTCGAGGTTAAAGATACTGAA 947
          || | | | |||    | |||  | |   | |||  || ||| || ||| | || | |
Db    482 GTGAATTACATCTTCTTTAAAGGCAAATGGAGGAGACCCCTTTGAAGTCAAGGACACCGAG 541
Qy    948 GAGGAAGATTTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATGAAAAGACTG 1007
          ||||| | ||  | || |||||| | | | |||  || || ||   |||||||  | |
Db    542 GAAGAGGACTTCCACGTGACCAGGTGACCGACGGCCTATGATGAGGCTGCCTATGATGAAGCGTTTA 601
Qy   1008 GGTATGTTCAATATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTATTGAAGTAT 1067
          ||||| || |  | || ||  ||    |  | |||  | |   | ||| | | | |||
Db    602 GGCATGTTTAACATCCAGCACTGTAAGAGCTGTCCAGCTGGTGCTGCTGATGAAATAC 661
Qy   1068 TTAGGTAACGCTACTGCTATTTTTTTTTTTACCAGACGAAGTTAAGCTTCAACATTTAGAG 1127
          || |||  | | | | |  || |||||||||  |  | || |  | ||| ||||||||
Db    662 CTGGGCAATGCCACCGCCATCTTCTTCTTCCTGCCTGATGAGGGAAACTACAGCACCTGGAA 721
Qy   1128 AATGAGTTGACTCATGACATTATTACTACTAAATT--TTTAGAGAACGAGGATCGTCGTAGCG 1185
          ||| || | |  |||| | | || || || |  |   |||||||| | | | | ||||
Db    722 AATGAAACTCACCCACGATATCATCACNCAGTTCCCTGGAAAATGAAGACAGAAGTCCTG 781
Qy   1186 CTTCTCTGCGACCTGCCAAAGTTAAGTATCACCGGTACTTACGACTTTGTTTTAG 1245
          |  | | | || |  || |||  | |  || |  |  || || || | | |  |
Db    782 CCACCTTACATTTACCCAAACTGTCCATTACTGGAAACTATGATCTGAAGAGCGTCTGG 841
Qy   1246 GCCAGTTAGGTATTACCAAAGTTTTTTCTAACGGTGCC 1283
          | |  | || | || |||   |||||| |  | ||
Db    842 GTCAACTGGGCATCACTAAGGTCTTCAGCAATGGGGGC 879
```

```
RESULT 11
BM924813
LOCUS       BM924813              1194 bp    mRNA        linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6653921 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5761267
            5', mRNA sequence.
ACCESSION   BM924813
VERSION     BM924813.1  GI:19375192
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1194)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov http://www.llnl.gov/
            Plate: LLAM12809  row: j  column: 20
```

```
Qy   900  TTTTTCAAGGGTAAGTGGGAACGTCCTTCGAGGTTAAAGATCTGAAGAGGAAGAGATTTT  959
          || || ||||||||||  |||||      || |||  ||||
Db   247  TTCTTTAAGGCAAATGGGAAGGACCCTTTGAAGTCAAGGACACCGAGGAAGGAGGACTTC  306

Qy   960  CATGTTGATCAAGTTACTACTGTCAAAGTTCCAATGATGAAAAGACTGGGTATGTTCAAT  1019
          || ||||| ||| ||| ||||| || ||  |||||
Db   307  CACGTGGACCAGGTGACCACCGTGAAGGTGCCTATGATGATGAAGGCCGTTTAGGCATGT  366

Qy  1020  ATTCAACATTGCAAAAAATTAAGTTCTTGGGTCTTATTAATGAAGTATTTAGGTAAACCT  1079
          |||||| ||    ||  ||||    ||| || ||||| || |||||
Db   367  ATCCAGCACTGTAAGAAGCTGTCCAGCTGGGTGGCTGCTGATGAAATACCTGGGCAATGCC  426

Qy  1080  ACTGCTATTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTTAGAGAATGAGTTGACT  1139
          ||| ||| ||  || |||  ||  ||  || || |||  ||||||  ||  ||||
Db   427  CACGATATCATCACCAAGTTCCTGGAAAATTGAAGACAGAAGGTCTGCCAGCTTACCATTTA  486

Qy  1140  CATGACATTATTACTAAATTTTTAGACAGAGAGATCGTCGTAGCGCTTCTCTGCACCTG  1199
          |||  ||  ||  |||||| || |||||  ||   || |||  || |||| |||  ||
Db   487  CACGATATCATCACCAAGTTCCTGAAAATTGAAGACAGAAGGTCTGCCAGCTTACCATTTA  546

Qy  1200  CCAAAGTTAAGTATCACCGGTACTTACGACTTAAAATCTGTTTTAGGCCAGTTAGGTATT  1259
          |||||  ||  ||| ||| | |||| || |||||||  |||
Db   547  CCCAAACTGTCCATTACTGAACCTATGATCTGAAGACGCCTCGGTCAACTGGGCATC  606

Qy  1260  ACCAAAGTTTTTTCTAACGGTGCCGATTTGAGTGGTGTTACTGAAGAAGTCCATTAAAA  1319
          || ||||| ||| || |||  ||| |||| || ||| ||| |||| ||  || ||||
Db   607  ACTAAGGTCTTCAGCAATGGGCTGACCTCTCCGGNGTCACAGAGGCACCCCTGAAG  666

Qy  1320  TTGAGTAAAGCTGTTCACAAAGCCGTCTTAACTATTGATGAAAAGGGTACCGAGGCCGCC  1379
          || ||||  ||| || |||| | ||  ||| |||| |||  |||  ||||
Db   667  CTCTCCAAGGCCGTGCATAAGGCTGTGTCGACCATCGACGAGAAAG-ACTGAAACTGCT  725

Qy  1380  GGGCGTATGTTCCTGGAAGCTATTCCAATGAGCATTCCACCAGAAGTTAAATTTAA--TA  1437
          ||| || ||| || |||  || ||  |||  |||  ||||
Db   726  GGAGCCATGTTTTTAGAGGGCCTACCCATGTTATCCCCCCGAGTCAAGGTTCACCAA  785

Qy  1438  AACCATTCGTTTTTCTGATGATCGAGCAGAACACTAAAGCCCATTGTTATGGGTAA  1495
          || |||| ||  || ||||   |||||
Db   786  ACCCTTTGTCTTCTTTAATGATTGGACCAAAATACCAAGTCTCCCCTCTTCCGGGGAA  843
```

RESULT 13
BQ646948
LOCUS   BQ646948   AGENCOURT_8302495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271313
DEFINITION AGENCOURT_8302495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271313
            5', mRNA sequence.
ACCESSION   BQ646948
VERSION   BQ646948.1 GI:21771120
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 891)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2447 row: n column: 18
            High quality sequence stop: 672.
FEATURES   Location/Qualifiers
     source   1..891
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6271313"

```
Db   829  GGCAACGGCCATCTTCTTCCGCCTGATGAGGGAGAAACTACAGCACTGGAAAA  881
```

RESULT 14
BQ648524
LOCUS   BQ648524   959 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION AGENCOURT_8297828 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269919
            5', mRNA sequence.
ACCESSION   BQ648524
VERSION   BQ648524.1 GI:21772696
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 959)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2444 row: d column: 16
            High quality sequence stop: 674.
FEATURES   Location/Qualifiers
     source   1..959
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6269919"
                     /clone_lib="NIH_MGC_100"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver;Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."

BASE COUNT   246 a   276 c   245 g   192 t
ORIGIN

Query Match   17.7%; Score 270.4; DB 14; Length 959;
Best Local Similarity   59.3%; Pred. No. 5.1e-59;
Matches   477; Conservative   0; Mismatches 326; Indels 1; Gaps 1;

Qy   510 ACTGCTTTCGCCATGTGAGTTAGGTACTAAAGCCGATACCCATGGACGAGATTTTAGAA 569
              | |||| || ||||| | || || | ||||| || |||| |||| | |||| |
Db   1 ACAGCCTTTGCAATGCTCTCCCTGGGACCAAGGCTGACACTCACGATGAAATCCTGGAG 60

Qy   570 GGTTTAAACTTTAATTTGACCGAAATCCCAGAAGCCAAATCACGAGGGTTTTCAAGAG 629
          || | |||| || ||| ||| | || ||| | ||| | ||| | | || |||||
Db   61 GGCCTGAATTTCAACCTCACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCCAGGAA 120

Qy   630 TTGTTGAGAACTTTGAATCAACCTGATTCTCAATTGCAATTAACTACTGGTAACGGTTTA 689
              | | || | | |||| |||| |||||| | |||| | | ||| | ||
Db   121 CTCTCCGTACCCTCACCAGCCAACCAGCCCAGCTGACCTCCAGCGCAATGGCCTG 180

Qy   690 TTTTTTTGTCTGAAGGTTAAAATTGGTTGACAAATTCCTGAGAGACGTCAAGAAACTATAT 749
           || || | |||| ||||||| | |||| ||||| | ||| || || ||||| | |
Db   181 TTCCTCAGCGAGGGCCTGAAGCTAGTGGATAAGTTTTTGGAGATGTTAAAAGTTGTAC 240

Qy   750 CATAGTGAGGCTTTTACCGTTAATTTGGTGATACTGAGGAAGCTAAAAAGCAAATTAAT 809
          ||| |||| | || | |||| | ||||| |||||| |||||||||||| |||||||||
Db   241 CACTCAGAAGCCTTCACTGTCAACTTCGGGGACACCAGAGGCCAAGAAGAGGCCAAGAAACAGATCAAC 300
```

```
BASE COUNT    229 a    247 c    226 g    189 t
ORIGIN

Query Match             17.8%;  Score 271.4;  DB 14;  Length 891;
Best Local Similarity   59.4%;  Pred. No. 2.7e-59;
Matches   495;  Conservative   0;  Mismatches 336;  Indels 2;  Gaps 2;

Qy   298  TGTCTGGTAAGTCCTGTGTTCCCCAGTCAAGGCCATGGAAGACCCTCAAGGCGACGCCG  357
Db    50  TGGCAGGCCGTGCTGCCGCTGGTCCCTGTCCCCGGCTGAGGGGAGATGCTG  109

Qy   358  CTCAAAAAACCGACACCAGTCATCACGACCAAGACCATCCGACTTTTAATAAAATTACTC  417
Db   110  CCCAGAAGACAGATACATCCCCATGATCAGGATCACCCAACCTTCAACAAGATCACCCC  169

Qy   418  CAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAATTAGCTCATCAAAGTAATTCTA  477
Db   170  CCAACCTGCTGAGTTCGCCCTTCAGCCTATACCGCCAGTCTGGCACACCAGTCCAACAGCA  229

Qy   478  CTAACATTTTTTTTAGTCCTGTTCTATTGCCACTGCTTTCGCCATGTTGAGTTTAGGTA  537
Db   230  CCAATATCTTCTTCTCTCCCCAGTGAGCATCGCTACAGCCTTTGCAATGCTCTCCCTGGGA  289

Qy   538  CTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTTAAACTTTAATTTGACCGAAATCC  597
Db   290  CCAAGGCTGACACTCACGATGAAATCCTGGAGGGGCCTGAATTTCAACCTCACGGAGATTC  349

Qy   598  CAGAAGCCCAAATTCACGAGGGTTTTCAAGAGTTGTTGAGAACTTTGAATCAACCTGATT  657
Db   350  CGGAGGCTCAGATCCATGAAGGCTTCCAGGAACTCTCCGTACCCTCCAACCAGCCAGACA  409

Qy   658  CTCAATTGCAATTAACTACTGGTAACGGTTTATTTTTTTGTCTGAAGGTTTAAAATTGGTTG  717
Db   410  GCCAGCTGACCTCCAGCGCAATGGCCGCCGTGTTCCTCAGCGAGGGCCTGAAGCTAGTGG  469

Qy   718  ACAAATTCCTAGAAGACGTCAAGAAACTATATCATAGTGAGGCTTTTACCGTTAATTTG  777
Db   470  ATAAGTTTTTGGAGATGTTAAAAGTTGTACCACTCAGAGAGCCTTCACTGTCAACTTCG  529

Qy   778  GTGATACTGAGGAAGCTAAAAAGCAAATTAATGATTATGTTGAGAAAGGCACCCAGGGTA  837
Db   530  GGGACACCGAAGAGGCCAAGAAGAAACAGATCAACGATTACGTGGAGAAGGGTACTCAAGGGA  589

Qy   838  AGATCGTTGACCTAGTTAAAGAATTAGATCGTGATACCGTCTCGACACTAGTTAACTATA  897
Db   590  AATTTGGATTTGGTCAAGGAGCTTGAGCAGAGACAGCAGTTTTGCTCTTGGTGAATTACA  649

Qy   898  TTTTTTTCAAGGGTAAGTGGGAAACGTCCTTTGACGGTTAAAGATACTGA-AGAGGAAGAT  956
Db   650  TCTTCTTAAAGGCAAATGGAGGACCCCTTTGAAGTCAAGGACACCGAGGGGAAGAGGAC  709

Qy   957  TTTCATGTTGATCAAGTTACTACTGGTCAAAGTCCAAATGATGAAAAGACTGGGTATGTTC  1016
Db   710  TTCCACGTGACCAGGTGACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCATGTTT  769

Qy  1017  AATATTCAACATTGCAAAAATTAAGTTCTTGGTCTTATTAATGAAGTATTTAGGTAAC  1076
Db   770  AACATCCAGCACTGTAAGAAGCTTGACCGTCCAGCTGGGCTGCTGACTGAA-TACCTGGGGCAT  828

Qy  1077  GCTACTGCTATTTTTTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTTAGAGAA  1129
```

```
            sfiI (ggcgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGGCCGAGGCGGCCGCGACATG-dT(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.85
            kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."

BASE COUNT      222 a      215 c      191 g      166 t        2 others
ORIGIN

Query Match              17.6%;   Score 268;   DB 12;   Length 796;
Best Local Similarity    60.0%;   Pred. No. 2e-58;
Matches 479;   Conservative 0;   Mismatches 317;   Indels 2;   Gaps 2;

Qy   379  ATCACGACCAGACCATCCGACTTTTAATAAAATTACTCCAAATTTAGCCGAATTTGCTT  438
               | ||||| || ||| |  |||| |||| || || || |||  |||||||||||||
Db     1  ACCATGATCAGGATCACCCAACCTTCAACAAGATCACCCCCAACCTGGCTGAGTTCGCCT   60

Qy   439  TTTCTTTGTATAGACAATTAGCTCATCAAAGTAATTCTACTAAGCATTTTTTAGTCCTG   498
                 ||  || ||| | |||  || || || ||| ||  ||| || |||| || || ||
Db    61  TCAGCCTATACCGCCAGCTGGCACACCAGTCCAACAGCACCAATATCTTCTCTCCCCAG   120

Qy   499  TTTTCATTGCCACTGCTTTCGCCATGTGAGTTTAGGTACTAAAGCCGATACCCATGACG   558
               |  || ||  | || | ||| || |   |  ||  ||  | |||  | || || ||
Db   121  TGAGCATCGCTACAGCCTTTGCAATGCTCTCCCGTGGGACCAAGGCTGACACTCACGATG  180

Qy   559  AGATTTTAGAAGGTTAAACTTTAATTGACCGAAATCCCAGAAGCCAAATTCACGAGG    618
               ||  || ||| |||  ||||| ||| |   |  ||  || |   || || || |||
Db   181  AAATCCTGGAGGGCCTGAATTTCAACCTCACGGAGATTCCGGAGGTCGAGATCATGATGAAG  240

Qy   619  GTTTTCAAGAGTTGTTGGAGAACTTTGAATCAACCTGATTCTCAATTGCAATTAACTACTG  678
               ||   ||| |  | || ||| ||    ||  ||  ||  | ||| || |||| || ||
Db   241  GCTTCCAGGAACTCCTCCATACCCCTCAACCCAGCCAGCCGCCAGCTCACGCTGACCACCG  300

Qy   679  GTAACGTTTATTTTGTCTCAGCGAGGTTAAAATTGGTTGACAAATTCCTAGAGACGTCA    738
               || ||  | ||  | ||||||||||  || | | ||| || || || |||||| ||
Db   301  GCAATGGCCGTGTTCCTCAGCGAGGGCCTGAAGCTAGTGGATAAGTTTNTGGAGGATGTTA  360

Qy   739  AGAAACTATATCATAGTGGAGGGCTTTTACCGTTAATTTGGTGATACTGAGGAAGCTAAAA  798
               |||| | ||  ||| || || || || ||  |   || || ||||| ||||| |||||
Db   361  AAAAGTTGTACCACTCAGAAGCCTTCACTGTCAAAGATGTTCGGGGAGATTCTGGTGGCCAAGA  420

Qy   799  AGCAAATTAATGATTAATGTTGAGAAAGGCACCCAGGGTAAGATCGTTGACCTAGTTAAAG  858
               || ||  |||| || | |||| ||| || ||  || ||| || || || ||  ||| |
Db   421  AACAGATCAACGATTACGTGGGAGAAGGTACTCAAGGAAAAT-GTGGATTTGGTCAAGG    479

Qy   859  AATTAGATCGTGATACCGTCTTCGCACTAGTAGTTAACTATATTTTTTCAAGGTAAGGTTGGG  918
               || ||  ||| || |  || || || ||| || |||  | |  ||  ||||  ||| |||
Db   480  AGCTTGACAGAGACACAGTTTTGCTCTGGTGAATTACATCTTCTTTAAAGGCAAATGGG    539

Qy   919  AACGTCCTTTCGAGGTTAAAGATACTGAAGAGGAAGAGATTTTCATGTTGATCAAGTTACTA  978
               | |||  |||  || |  || |  ||| || || ||  || |  | ||| ||  || || |
Db   540  AGAGACCCTTTGAGTCAAGGACACCGAGGAGAGGAGACTTCCACGTGGACCAGGTGACCA    599

Qy   979  CTGTCAAAGTTCAATGATGAAAAGACTGGGTATGTTCAATATTCAATATTCAACATTGCAAAAAAT  1038
               | ||  |  |||  ||| ||| || || || || || || || || | ||| ||  |||||
Db   600  CCGTGAAGGTGCCTATGATGAAGCGTTTAGCCATGTTTAACATCCAGCACTGTAAGAGC    659

Qy  1039  TAAGTTCTTGGGTCTTATTAATGAAGTATTAGGTAACGCTACTACTGCTATTTTTTTTTTTAC  1098
               | |  |  || ||  | || || || ||  | ||  || || || || || || ||  | |
Db   660  TGTCCAGCTGGGTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCGNCATCTTCCTGC    719

Qy  1099  CAGACGAAGGTAAGCTTCAACATTTAGAGAATGAGTTGACTCAGCATTATTACTACTAAAT  1158
               ||| || || || || || ||  | || || || || ||  ||||| || || || ||
Db   720  CTGATGAGGGGAAATTACAGCACCTGGAAAATACCTGGCCATGGCAATGCCGGCACCGNCANCAGT  778

Qy   810  GATTATGTTGAGAAAGGCACCCAGGGTAAGATCGTTGACCTAGTTAATTAGATCGT      869
               |||| |  ||  | | |||| ||||| | |  || |  ||||    ||  |||||
Db   301  GATTACGTGGGAAGAGGGTACTCAAGGAGAAATTGTGGATTTGGTCAAGGAGCTTGACAGA   360

Qy   870  GATACCGTCTTCGCACTAGTTAACTATATTTTTCAAGGTAAGGTAGGTGGAACGTCCTTTC   929
               || ||  || || || ||||   || |||  || |  || || |||  ||| ||  ||
Db   361  GACACAGTTTTGCTCTGGTGAATTACATCTTCTTTAAAGGCAAATGGGAGAGCACCCTTT   420

Qy   930  GAGGTTAAAGATACTGAAGAGAGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTT   989
               || ||  || || || ||||| || || || || | ||  ||||| ||  || || ||||
Db   421  GAAGTCAAGGACACCGAGGAGAGGAGACTTCCACGTGGACCAGTCGACCACCGTGAAGGTG  480

Qy   990  CCAATGATGAAAAGACTGGGTATGTTCAAATATTCAACATTGCAAAAAATTAAGTTCTTGG  1049
               || ||||| ||  | || || ||||| ||  |||||||| |   || |||| || ||
Db   481  CCTATGATGAAGCGTTTAGCATGTTTAACATCCAGCACTGTAAGAGCGTGTCCAGCTGG   540

Qy  1050  GTCTTATTAATGAAGTATTAGGTAACGCTACTACTGCTATTTTTTTTTTACCAGACGAAGGT  1109
               || || || || || ||  | ||  || || || || || || || || | || ||||
Db   541  GTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCATCTTCTTCCTGCCTGATGAGGGG   600

Qy  1110  AAGCTTCAACATTTAGAGAATGAGTTCACTCAGCATTATTACTAACATTTTTAGAGAAC   1169
               || || || ||  | || ||| ||||||  ||| | ||   |  |  ||   ||| ||
Db   601  AAACTACAGCACCTGGAAAATGAACTCACCCACGATATCATCACCAAGTTCCTGAAAAT    660

Qy  1170  GAGGATCGTCGTAGCGCTTCTCTGCACCTGCCAAAGTTAAGTATCACCGGTACTTACGA-  1228
               || || || |||| ||  | || || || ||  |  || || || || |  || ||| |
Db   661  TGAGCATCGCTACAGCCTTTGCAATGCTCTCCCGTGGGACCAAGGCTGACAGTCACGATG   720

Qy  1229  CTTAAAATCTGTTTTAGGCCAGTTAGGTATTACCAAAGTTTTTCTAAGCTGCCGATTT  1288
               || ||  | || ||  | || || || || ||  || ||||  |  ||| || || |
Db   721  CTGAAGAGGCGTCCTGGGTCAACTGGCATCACTCAAGGGCTTCAGCCATGGGCTGACT   780

Qy  1289  GAGTGGTGTTACTGAAGAAGCTCC  1312
               |  ||  || |||| || || ||
Db   781  CCCGGGGGTCACAGAGGAACGCC    804
```

```
RESULT 15
BG567260
LOCUS       BG567260                 796 bp    mRNA    linear    EST 10-APR-2001
DEFINITION  602589741F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723498 5',
            mRNA sequence.
ACCESSION   BG567260
VERSION     BG567260.1  GI:13574913
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 796)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1583  row: b  column: 11
            High quality sequence stop: 730.
FEATURES             Location/Qualifiers
     source          1..796
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4723498"
                     /clone_lib="NIH_MGC_76"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
```

Mon Dec 9 12:50:53 2002

us-10-025-514-7.rst

Search completed: December 7, 2002, 02:06:12
Job time : 1545.5 secs